

TEMI E ANALISI

Dati aperti e ricerca scientifica: aspetti gestionali e normativi nel contesto dell'e-science

di Maria Cassella

Because good research needs good data
(Digital Curation Centre)

Nel settore pubblico, nella ricerca, in ambito scientifico, ma anche nel campo delle scienze umane e sociali, vengono prodotte giornalmente grandi masse di dati in formato digitale che necessitano di essere analizzate, gestite e conservate in maniera adeguata per trasformarsi in informazioni strutturate e consentire il loro riutilizzo futuro.

Dati ed informazioni non sono tra loro sinonimi.

I dati sono numeri, caratteri, immagini, formule, che designano l'attributo di un fenomeno.

«Dal punto di vista del linguaggio giuridico “dati” ha una portata semantica più ristretta e si riferisce appunto solo alle singole e isolate informazioni, non organizzate e non elaborate dall'ingegno umano»¹.

Per trasformarsi in informazione i dati grezzi devono essere strutturati: «l'informazione può essere definita come lo scarto di qualità dal dato grezzo, il quale deve anche essere strutturato secondo una sintassi che lo renda interpretabile»².

Un'altra distinzione che ci sembra utile proporre per fugare dubbi o fraintendimenti è quella tra dataset, banche dati (*database*) e metadati.

I dataset sono insiemi di dati omogenei, strutturati³ prodotti da un unico soggetto o da soggetti diversi, pubblici o privati, nello svolgimento di un'attività o di

MARIA CASSELLA, Università degli studi di Torino, Sistema bibliotecario di ateneo, via Verdi, 8, 10124 Torino, e-mail maria.cassella@unito.it.

Ultima consultazione siti web: dicembre 2013.

1 Simone Aliprandi, *Open Licensing e banche dati*, «Informatica e diritto», (2011), n.1/2, p. 25-44, accessibile alla URL <http://www.aliprandi.org/pub/aliprandi_ied_database.pdf>.

2 Antonella De Robbio – Silvia Giacomazzi, *Dati aperti con LODE*, «Bibliotime», 14(2011), n. 2, <<http://www.aib.it/aib/sezioni/emr/bibttime/num-xiv-2/derobbio.htmA>>. Sui dati aperti si legga anche: Antonella De Robbio, *Dati aperti nella Pubblica Amministrazione tra crescita e trasparenza*, «Digitalia», (2013), n. 1, <<http://digitalia.sbn.it/article/view/718A>>.

3 Generalmente in file tabulari (CSV, XML, CALC ecc.).

un processo di ricerca. In quanto strutturati possono essere facilmente interrogabili e interpretabili da applicazioni di tipo diverso; le banche dati sono, invece, «archivi di dati collegati secondo un modello logico e in modo tale da consentire la gestione dei dati stessi da parte di particolari applicazioni software dedicate»⁴.

I metadati sono, infine, dati strutturati sui dati, ne garantiscono la scoperta, la consultazione, il riuso, la conservazione a medio e lungo termine. Rispetto ai dati i metadati forniscono una serie di informazioni quali, ad esempio, le informazioni sul creatore, sulle metodologie di acquisizione, sulla data di creazione ecc.; ne descrivono i dettagli tecnici, le metodologie di selezione e di analisi, i diritti di riutilizzo.

In tempi recenti abbiamo assistito ad un fiorire di riflessioni sull'apertura in rete dei dati⁵ pubblici (intesi come dati di possesso delle pubbliche amministrazioni) e dei dati della ricerca.

Il tema matura in un contesto di forte innovazione culturale, tecnologica, legale e sociale: da un lato sfrutta l'interesse dell'Unione Europea teso ad incrementare lo sfruttamento economico dei dati delle amministrazioni pubbliche⁶ e le iniziative pubbliche e private volte ad aumentare la creazione di contenuti digitali (ad esempio Europea); dall'altro si inserisce nella rivendicazione di un accesso ampio e democratico ai prodotti della ricerca scientifica e, quindi, si innesta nella scia del movimento Open Access.

In questo articolo ci occuperemo solo marginalmente dei dati prodotti dalle pubbliche amministrazioni. Nella prospettiva dell'Open Government (o *e-government*) il tema dei dati aperti tocca la ricerca così come l'informazione del settore pubblico: gli Open data sono, infatti, una delle condizioni necessarie per raggiungere gli obiettivi di trasparenza della P.A., di diffusione dell'informazione e di crescita economica di un Paese.

Rifletteremo, invece, prevalentemente sul tema dei dati della ricerca che, gestionalmente, stanno aprendo nuovi scenari per le comunità scientifiche nonché per le biblioteche accademiche e degli enti di ricerca, ponendo nuove sfide ed interrogativi. Partendo da alcune definizioni di "dati aperti" cercheremo di capire come, dove e quando i dati possono essere messi liberamente a disposizione in rete, a quali condizioni, con quali licenze, chi è responsabile della loro gestione e conservazione. L'ultimo capitolo è dedicato al tema dei dati aperti come nuovo paradigma per le biblioteche accademiche e di ricerca.

Dati della ricerca e dati aperti: alcune definizioni

I dati della ricerca sono modellati sulle comunità scientifiche che li producono utilizzando approcci e tecnologie differenti: vista la loro enorme varietà, origine, forma, quantità, ogni comunità di ricerca può adottare una propria definizione.

Nella letteratura sul tema emergono, quindi, molteplici definizioni di dati della ricerca. Nel volume *OECD Principles and guidelines for access to research data from public funding* pubblicato nel 2007 dall'Organization for Economic Cooperation and Development (OECD)⁷ i dati della ricerca vengono definiti come: «record fattuali (dati numerici, record testuali, immagini e suoni) utilizzati come fonti primarie per la

⁴ Definizione tratta dalla voce "database" di Wikipedia Italia, <<http://it.wikipedia.org/wiki/Database>>.

⁵ D'ora in avanti il termine "dati" verrà utilizzato nell'articolo come sinonimo di "dataset".

⁶ Si pensi al valore dei dati pubblici per le imprese che possono utilizzarli per sviluppare ed indirizzare strategie commerciali e di marketing.

⁷ OECD, *Principles and guidelines for access to research data from public funding*, April 2007, <<http://www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>>.

ricerca scientifica. Tali fonti sono comunemente accettate nella comunità scientifica come necessarie per convalidare i risultati della ricerca».

Più stimolante è distinguere le diverse tipologie di dati della ricerca e cercare di suggerire una loro classificazione. Paola Gargiulo (2013) propone la seguente categorizzazione ritagliata sulla diversa origine dei dati:

- dati risultati da osservazioni;
- dati sperimentali (dati di laboratorio riproducibili, ma costosi);
- dati generati da simulazioni (per esempio modelli climatici, modelli economici);
- dati derivati o compilati (per esempio i dati estratti tramite *data/text mining*, i modelli 3D; riproducibili ma costosi);
- dati di riferimento o dati di canone (per esempio le banche dati di sequenze di geni, le strutture chimiche, i portali con dati spaziali)⁸.

Nell'ambito della ricerca scientifica si discute da alcuni anni di interoperabilità e di apertura in rete dei dati della ricerca, in quanto si ritiene che siano strategici ai fini dell'avanzamento dell'*e-science*,⁹ ma anche per la crescita del settore privato, che può beneficiare considerevolmente del trasferimento di conoscenze. Sotto un profilo più ampio e generale l'accesso alle ricerche scientifiche e ai dati associati favorisce lo sviluppo economico di un Paese e ne accresce il benessere sociale.

Nella ricerca scientifica l'accesso ai dati comporta una serie di benefici:

- rafforza l'indagine scientifica;
- incoraggia la diversità di analisi e di opinioni;
- promuove la ricerca cross e multidisciplinare;
- rende possibile testare nuove ipotesi di ricerca e metodi di analisi alternativi;
- facilita l'istruzione e la formazione dei giovani ricercatori;
- consente di esplorare ambiti di ricerca nuovi e non precedentemente oggetto di altre ricerche;
- consente la creazione di nuovi dataset mediante la combinazione di dati provenienti da fonti diverse.

In campo scientifico l'attenzione sul tema dei dati aperti della ricerca scientifica è stata posta a partire dal 2006 dal movimento Open Data che annovera tra i suoi fondatori ed esponenti di maggior spicco Peter Murray-Rust, biochimico dell'Unilever Centre dell'Università di Cambridge.

Murray-Rust definisce i dati aperti come «dati che possono essere utilizzati, studiati e modificati senza restrizioni e che possono essere copiati e ridistribuiti in una forma modificata o non modificata senza restrizioni o con l'unica restrizione volta a garantire che altri contenitori possano farne lo stesso uso»¹⁰.

Lo studioso traccia quindi una significativa distinzione tra i dati prodotti nella "*small science*" e quelli prodotti nella "*big science*", evidenziando come il trattamen-

8 Cfr. Paola Gargiulo, *La gestione dei dati della ricerca e il ruolo delle biblioteche: quali sfide e quali prospettive*, Ciclo di conferenze: "Biblioteche, libri e documenti: dalla informazione alla conoscenza", a.a. 2012-2013, Roma: La Sapienza, 19 aprile 2013, <<http://www.slideshare.net/PaolaGar/dati-della-ricerca-e-bibliotecari-20480165>>.

9 Per *e-science* si intende quella scienza che utilizza sistemi di calcolo molto potenti e performanti in rete e si basa su un'infrastruttura condivisa tra le diverse comunità di ricerca.

10 Peter Murray-Rust, *Open data in science*, «Serials review», 34 (2008), n. 1, p. 52-64. Il preprint dell'articolo è disponibile su Nature Proceedings alla URL: <<http://proceedings.nature.com/documents/1526/version/1>>. La traduzione del testo in italiano è mia.

to dei dati e la sensibilità verso la loro condivisione dipenda in gran parte dalle prassi delle comunità di ricerca: «si riconosce che la pratica nella pubblicazione e riutilizzo dei dati varia notevolmente tra le diverse discipline. Alcune di queste, come le bioscienze, hanno una lunga tradizione nel richiedere dati da pubblicare e aggregare in banche dati a finanziamento pubblico. Altre discipline, o “scienze grandi”, hanno ben sviluppato politiche del riutilizzo dei dati e richiedono questi ultimi da telescopi, satelliti, acceleratori di particelle, sorgenti di neutroni, ecc., affinché diventino universalmente disponibili per il riuso. In queste aree le norme sono spesso sufficienti per la pratica dell’accesso aperto ai dati. Nelle “scienze piccole” invece l’unità di ricerca è il laboratorio o l’individuo (con il termine “scienze piccole” non si intende riflettere l’importanza della disciplina che può essere numericamente vasta). Queste discipline tipicamente sfociano in molte pubblicazioni indipendenti che riportano esperimenti individuali (*hypopublication* è un neologismo che esprime la natura disarticolata di tali informazioni)»¹¹.

Un’indagine condotta in Gran Bretagna nel maggio 2008 tra i ricercatori di quattro università (università di Bristol, di Leeds, di Leicester e di Oxford)¹² ha messo in evidenza come nelle pratiche di condivisione dei dati i più attivi siano gli studiosi di scienze sociali (45%), seguiti dagli umanisti (37%) e, quindi, dagli studiosi di area scientifica (22%).

Si ispira alla consolidata definizione di accesso aperto della Budapest Open Access Initiative (febbraio 2002)¹³ la definizione di “dati aperti” contenuta nei Principi Panton (*Panton Principles*)¹⁴ concepiti nel 2009 e pubblicati nel 2010 dall’Open Knowledge Foundation Working Group on Open Data in Science: «con dati aperti nella scienza intendiamo la possibilità per ogni utente di scaricare, copiare, analizzare, rielaborare, inviare a software o usare i dati per altri scopi gratuitamente senza barriere finanziarie, legali o tecniche se non quella dell’accesso alla stessa rete Internet».

Una terza definizione di “dati aperti” ampiamente condivisa e adottata è quella dell’Open Knowledge Foundation (OKFN)¹⁵ che li descrive come «dati che possono essere liberamente utilizzati, riutilizzati e ridistribuiti da chiunque alla condizione che venga specificata l’attribuzione e che venga adottato il principio del “condividi allo stesso modo” (*sharealike*)»¹⁶.

L’OKFN specifica, quindi, i requisiti per l’apertura dei dati e dei contenuti: *Disponibilità e Accesso*: i dati devono essere disponibili nella loro interezza, in un formato conveniente e modificabile. I costi di riproduzione devono essere ragionevoli; *Riuso e Distribuzione*: i dati devono essere resi disponibili in modo tale da consentirne il riuso e la ridistribuzione. Devono essere comprensibili alle macchine (*machine-readable*);

11 Peter Murray-Rust, *Open data in science* cit. La traduzione del testo in italiano è di De Robbio – Giacomazzi.

12 Neil Beagrie – Robert Beagrie – Ian Rowlands, *Research data preservation and access: the views of researchers*, ARIADNE, 2009, n. 60, <<http://www.ariadne.ac.uk/issue60/beagrie-et-al/>>.

13 <<http://www.budapestopenaccessinitiative.org/>>.

14 <<http://pantonprinciples.org/>>.

15 L’Open Knowledge Foundation è un’organizzazione no-profit fondata nel 2004 che ha tra i suoi obiettivi quello di promuovere l’utilizzo e la condivisione della conoscenza in rete, in tutte le sue forme.

16 <<http://okfn.org/opendata/#sthash.oPxvr4vf.dpuf>>.

Partecipazione Universale: ognuno deve essere libero di utilizzare, riutilizzare, ridistribuire i dati senza discriminazioni di settore, persone o gruppi¹⁷.

Altri requisiti per rendere efficace l'apertura dei dati in rete sono: una chiara ed efficace comunicazione per pubblici di tipo diverso (specialisti e non) ed una dichiarazione circa la validità dei dati stessi.

L'accesso ai dati della ricerca è, dunque, un tema complesso ed è condizionato dall'interazione di diversi fattori:

- *il fattore tecnologico*: l'accesso ai dati richiede una robusta infrastruttura tecnologica, basata su standard aperti che ne garantiscano l'interoperabilità. Quest'ultima è il cardine tecnologico e concettuale dell'apertura dei dati in rete. Per migliorare l'interoperabilità tecnologica e semantica dei dati archiviati in infrastrutture di rete di vario tipo è fondamentale che i dataset vengano esposti in modo strutturato, secondo formati aperti, siano arricchiti da metadati di qualità e dotati di identificativi persistenti. La mancanza di standardizzazione e la scarsa qualità dei metadati associati ai dati pregiudica fortemente la possibilità di accedere e di riutilizzare i dati.

Tim Berners-Lee, padre del World Wide Web e ideatore della parabola del web semantico,¹⁸ propone una scala crescente di valori e di complessità (il cosiddetto modello delle stelle) per aprire i dati in rete e renderli riutilizzabili:¹⁹

- mettete a disposizione i vostri dati sul web (in qualsiasi formato “così come sono”) ma con una licenza open;
- metteteli a disposizione come dati strutturati (per esempio in formato Excel, anziché come immagini scannerizzate di una tabella);
- utilizzate formati non proprietari/aperti (ad esempio CSV anziché Excel);
- usate gli standard aperti del W3C (RDF e SPARQL) per identificare gli oggetti, così che le persone possano creare puntatori alle vostre risorse;
- collegate i vostri dati ai dati di altre persone per permettere di capire in modo automatico il contesto (cioè quale è il loro contenuto).

La creazione di dataset aperti e collegati in rete (Linked Open Data) è la condizione più idonea per permettere alle macchine di costruire una semantica del web e realizzare servizi a valore aggiunto per gli utenti finali. Il dato in relazione tende ad assumere, quindi, un valore economico oltre che sociale e culturale.

- *il fattore manageriale*: la diversità e frammentazione delle comunità scientifiche suggeriscono un approccio alla gestione dei dati ritagliato sui bisogni delle singole comunità di ricerca. E' complesso in questa sede entrare nel merito della questione se sia preferibile un approccio istituzionale o disciplinare, locale, nazionale o internazionale nella gestione dei dati della ricerca. Si può, tuttavia, concordare sul fatto che, diversamente dal caso dei dati del settore pubblico che beneficiano dell'esi-

17 Dal sito dell'OKFN. La traduzione in italiano è mia.

18 L'idea di web semantico viene lanciata nel 2001 in un articolo di Tim Berners-Lee – James Hendler – Ora Lassila, *The semantic web*, «Scientific American», May 2001, <<http://www.scientificamerican.com/article.cfm?id=the-semantic-web>>. Sul web semantico si legga anche: Tim Berners-Lee, *Weaving the web: the original design and ultimate destiny of the World Wide Web*, Collins, 2000. Traduzione italiana: Tim Berners-Lee, *L'architettura del nuovo Web*, Milano: Feltrinelli, 2001. Sul web semantico è ricco di informazioni il sito del W3C: <<http://www.w3.org/standards/semanticweb/>>.

19 Il modello delle stelle proposto da Tim Berners-Lee è sul sito del W3C alla URL: <<http://www.w3.org/DesignIssues/LinkedData.html>>.

stenza di alcuni portali nazionali,²⁰ molteplici approcci gestionali sono possibili per i dati della ricerca, dal momento che le modalità di gestione e conservazione dipendono in gran parte dall'infrastruttura esistente, dalle esigenze delle comunità scientifiche e dai canali di comunicazione che queste ultime adottano;

– *il fattore economico*: il mantenimento di un'infrastruttura dedicata ai dati della ricerca, a qualsivoglia livello, richiede un budget dedicato e una precisa pianificazione economica che preveda, tra l'altro, un'assunzione di responsabilità da parte dell'ente produttore a garanzia di una sostenibilità nel lungo termine nella gestione e conservazione dei dati.

A tal fine si renderà necessario valutare opportunamente i costi-benefici del rendere accessibili i dati in rete.

Clyfford Lynch²¹ sottolinea come la messa a punto di una corretta strategia di gestione e conservazione dei dati prodotti dalla ricerca scientifica imponga sempre una valutazione delle diverse tipologie di dati, distinguendo, in base alle specificità disciplinari, tra dati che possono essere facilmente riprodotti e dati che non è possibile riprodurre se non a mezzo di investimenti elevati.

– *il fattore legale e l'assetto politico*: le leggi nazionali, in particolare quelle sulla tutela della proprietà intellettuale e della privacy, e gli accordi internazionali hanno un impatto enorme sulla possibilità di riutilizzare i dati della ricerca. In modo particolare l'accesso ai dati prodotti nell'ambito della cosiddetta “*big science*” può risultare fortemente condizionato dagli accordi internazionali e dalle *partnership* pubblico-private che vengono realizzate per finanziare progetti di ricerca di ampia portata.

– *il fattore culturale e comportamentale*: le diverse comunità di ricerca hanno un comportamento non omogeneo verso l'apertura dei dati in rete. Non è solo un problema culturale. Un serio ostacolo alla diffusione dei dati in rete è la mancanza di un sistema di validazione dei dati stessi. La messa a punto di metodologie per la certificazione dei dataset,²² l'inserimento dei dataset tra i prodotti della ricerca scientifica e la diffusione di nuove metriche di valutazione potrebbero in futuro modificare il comportamento delle comunità scientifiche meno propense a condividere i propri dati²³.

Dati aperti e licenze di uso

Le licenze CC

Un aspetto fondamentale degli Open data, ripreso e sottolineato in numerose definizioni, è la possibilità di riutilizzo dei dati. Infatti, diversamente dal caso della disseminazione in rete di articoli e monografie scientifiche, la disseminazione dei

20 Alcune considerazioni su un'infrastruttura nazionale dei dati del settore pubblico sono contenute nel report pubblicato da Science-Metrix a cura di Aurore Nicol - Julie Caruso - Éric Archambault, *Open data access policies and strategies in the European research area and beyond*, August 2013, <http://www.science-metrix.com/pdf/SM_EC_OA_Data.pdf>.

21 Cfr. il post Clifford Lynch, *scholarly works, big data and libraries*, del 23 settembre 2011 sul blog “Digitization 101” di Jill Hurst-Whal: <<http://hurstassociates.blogspot.it/2011/09/clifford-lynch-scholarly-works-big-data.html>>.

22 Nel campo della termo-fisica, ad esempio, un servizio di validazione dei dati per 5 riviste tematiche è offerto dal US National Institute of Standards and Technology.

23 È interessante evidenziare il caso di alcune riviste nate allo scopo precipuo di pubblicare in rete i dataset. Tra queste: *Acta Crystallographica E*, *Ecological Archives*, *GigaScience* e *International Journal of Robotics Research*.

dati, così come accade anche nel caso del software libero,²⁴ assume un valore reale, scientifico ed economico, solo se i dati possono essere concretamente riutilizzati dalle comunità di ricerca. Per consentire il riuso dei dati è fortemente consigliabile che chi possiede i diritti di proprietà intellettuale sugli stessi espliciti chiaramente quali forme di riutilizzo vengono consentite adottando lo strumento legale delle licenze. La pubblicazione in rete di dati non associati ad alcuna licenza determina, infatti, una situazione di incertezza che è la peggior nemica del riutilizzo.

Le licenze, invece, creano un «un diritto personale – o un insieme di diritti – per l'utente – o per un insieme di utenti – che accede a un dato – o a un insieme di dati – controllato da chi ne detiene la titolarità»²⁵.

L'uso delle licenze costituisce una buona pratica: riduce i costi di transazione, ovvero i costi monetari, le perdite di tempo e l'incertezza, sia per i detentori di dati, sia per i riutilizzatori.

Numerose licenze consentono l'accessibilità in rete ai dataset e disciplinano il loro riutilizzo. In linea teorica le licenze maggiormente idonee al riutilizzo dei dati in un contesto aperto e interoperabile sono le licenze di pubblico dominio. Per i dati le licenze più idonee sono quindi la CCo 1.0, che fa parte del noto gruppo di licenze Creative Commons e l'Open Data Commons Public Domain Dedication and License (PDDL) 1.0²⁶.

Il progetto Creative Commons nasce nel 2001 negli Stati Uniti per iniziativa di alcuni illustri giuristi. Ha concepito sei diversi tipi di licenze, dalla più permissiva – la licenza “attribuzione” – a quella meno permissiva, la licenza “non commerciale”, più una licenza di pubblico dominio: la CCo²⁷.

Le sei licenze CC si collocano a metà strada tra il regime del copyright (“tutti i diritti riservati”) e il pubblico dominio (“nessun diritto riservato”) e si basano, quindi, sul principio “alcuni diritti riservati”; è bene sottolineare il concetto che i diritti in capo all'autore scaturiscono dalla legge e non dalle licenze, il cui compito è, invece, quello di disciplinare l'utilizzo da parte di terzi di alcuni diritti sull'opera.

La CCo è la licenza della rinuncia a tutti i diritti nella misura permessa dalla legge applicabile;²⁸ è la licenza del gruppo CC più idonea a “licenziare” i dataset e i database; di fatto è una delle licenze consigliate dai Principi per i dati bibliografici aperti (*Principles for Open Bibliographic Data*)²⁹ e dai Principi Pantan.

Se il database non è licenziabile sotto la licenza CCo o il titolare dei diritti sceglie di non utilizzare una licenza di pubblico dominio è, comunque, possibile, anche se

24 Non è un caso che il principio 3 dell'Open Source Definition reciti: «la licenza deve consentire modifiche e opere derivate e permettere che il software venga distribuito alle stesse condizioni di licenza del software originale».

25 Antonella De Robbio, *OL4OD: licenze aperte per dati aperti*, «JLIS.it», 2 (2011), n. 2, <<http://leo.cilea.it/index.php/jlis/article/view/4695>>.

26 Una terza licenza di pubblico dominio è il Public Domain Mark 1.0 (Marchio di Pubblico Dominio) rilasciato nel 2010 da Creative Commons e dedicato a quelle opere che alla scadenza della tutela dei diritti rientrano nel pubblico dominio.

27 In realtà esiste anche una settima licenza CC: la CCPlus.

28 In Italia, secondo la vigente legge sul diritto di autore 633/1941 e successive modificazioni, i diritti morali sull'opera sono inalienabili.

29 <<http://openbiblio.net/principles/it/>>.

sconsigliabile per le ragioni di interoperabilità che esamineremo in seguito, adottare una delle sei licenze CC standard³⁰.

In Europa il quadro giuridico relativo ai database è reso più complesso dall'esistenza del cosiddetto diritto *sui generis*³¹ (Direttiva 96/9/EC dell'11 Marzo 1996) che, per la durata di 15 anni, protegge i database oggetto di un investimento rilevante sotto il profilo qualitativo e quantitativo.

Il trattamento del diritto *sui generis* da parte delle licenze CC è stato fino alla loro penultima versione (la 3.0) parziale e poco approfondito.

Per i Paesi dell'Unione Europea, infatti, le licenze CC 3.0 contenevano un'unica clausola relativa al diritto *sui generis* che consisteva nella rinuncia (*waiver*) al diritto stesso: «qualora l'Opera concessa in licenza includa una o più banche dati sulle quali il Licenziante è titolare di un diritto *sui generis* ai sensi delle norme nazionali di attuazione della Direttiva 96/9/CE sulle banche dati, il Licenziante rinuncia a far valere il diritto corrispondente».

Il trattamento del diritto *sui generis* da parte delle licenze CC 3.0 è stato oggetto di numerose critiche³².

A partire dalla versione 4.0, pubblicata a novembre 2013 mentre questo articolo veniva sottoposto al processo di revisione, le licenze CC regolano anche la distribuzione dei database e il loro riutilizzo.

Le principali novità delle nuove licenze sono:

- l'abbandono del sistema di adattamento agli ordinamenti giuridici nazionali (il cosiddetto sistema di *porting*)³³. A partire dalla versione 4.0, infatti, si fa riferimento ad un'unica versione internazionale delle licenze, il cui testo viene tradotto linguisticamente, ma non subisce più adattamenti alle legislazioni nazionali;
- il trattamento del diritto *sui generis* che consentirà l'utilizzo di tutte le licenze CC anche per i database.

Le licenze Open Data Commons

In quanto concepite in seno all'ordinamento giuridico statunitense le primissime versioni delle licenze CC (1.0, 2.0, 2.5) non contemplavano neppure la possibilità di rinunciare alla tutela del diritto *sui generis*. Per colmare questo vuoto nel 2006 la società informatica Talis, con sede legale a Birmingham e focus aziendale sul web semantico, studiò e pubblicò la prima licenza specificatamente dedicata ai dati e, successivamente, finanzia nel 2008 la nascita del progetto indipendente Open Data Commons (ODC),³⁴ allo scopo di realizzare strumenti legali per la condivisione in rete dei database.

30 Sul sito Creative Commons è disponibile un elenco di istituzioni che hanno adottato licenze CC per i dataset: <http://wiki.creativecommons.org/Data_and_CC_licenses>.

31 Recepita in Italia con d.leg. 6 maggio 1999, n. 169. Ai fini della Direttiva 96/9 per "banca di dati" si intende una raccolta di opere, dati o altri elementi indipendenti sistematicamente o metodicamente disposti ed individualmente accessibili grazie a mezzi elettronici o in altro modo.

32 Cfr. Lucie Guibault, *Licensing Research Data under Open Access conditions*. In: D. Beldiman (ed.), *Information and Knowledge: 21st century challenges in intellectual property and knowledge governance*, Cheltenham: Edward Elgar, in corso di pubblicazione, accessibile alla URL: <http://www.ivir.nl/publications/guibault/Open_Research_Data.pdf>.

33 Nel concetto di *porting* si racchiudeva una traduzione delle licenze nelle varie lingue, ma anche un contestuale adattamento delle clausole ai diversi ordinamenti giuridici.

34 Si tratta di un progetto britannico nato in seno all'università di Edimburgo per iniziativa del giurista Jordan Hatcher.

Nel 2008 il progetto ODC concepì e pubblicò l'Open Data Commons Public Domain Dedication and License (PDDL) per rispondere all'esigenza di creare una licenza di pubblico dominio utile a rilasciare in pubblico dominio un database o il suo contenuto o entrambi. Si legge, infatti, nel preambolo della licenza: «molti database sono coperti da copyright. Alcune giurisdizioni, soprattutto in Europa, hanno specifici diritti per la protezione dei database: ad esempio il diritto *sui generis*. Questi tipi di diritti, così come altri diritti legali utilizzati per proteggere i database e i dati, possono creare incertezza o difficoltà pratica per coloro che desiderano condividere le banche dati e i dati sottostanti conservando, nello stesso tempo, un numero limitato di diritti secondo il modello “alcuni diritti riservati”, così come illustrato nel *Science Commons Protocol for Implementing Open Data Access*. Come risultato, questa rinuncia e licenza cerca nella misura più ampia possibile di eliminare o licenziare appieno tutti i diritti che coprono questo database e i dati sottostanti».

Attualmente la licenza PDDL 1.0³⁵ è mantenuta dall'Open Knowledge Foundation. Il più noto utilizzatore di licenza PDDL è stato il progetto cartografico Open Street Map³⁶.

Del pacchetto di licenze OKFN fanno parte anche l'Open Database License (ODbL) e l'Open Data Commons Attribution (ODC-BY). La prima è una licenza virale di tipo “condividi allo stesso modo” (*share-alike*) concepita appositamente per le banche dati.

“Licenzia” unicamente i diritti relativi al database; qualora si tratti di un database contenente opere creative, per garantire un libero utilizzo dell'intera opera è opportuno applicare un'altra licenza alle opere contenute nel database stesso³⁷.

L'Open Data Commons Attribution (ODC-BY)³⁸ è stata rilasciata nel 2010; è una licenza di tipo “attribuzione”. Tra i diritti concessi la licenza include: a) estrazione e riutilizzo dell'intero database o di parti sostanziali di esso; b) creazione di database derivativi; c) creazione di database collettivi; d) creazione di riproduzioni temporanee o permanenti con ogni mezzo e in ogni forma, in tutto o in parte, inclusi database derivati e parti di database collettivi; e) distribuzione, comunicazione, esposizione, prestito, disponibilità e rappresentazione in pubblico con ogni mezzo e in ogni forma, in tutto o in parte, inclusi database derivati e parti di database collettivi.

La ODC-BY è stata raccomandata nel 2012 da OCLC alle biblioteche che riversano i dati nel catalogo WorldCat per consentirne il riuso. La scelta di una licenza “attribuzione” è allineata alla policy sui diritti e le responsabilità nella gestione di WorldCat³⁹ pubblicata da OCLC nel 2010 ed è compatibile con la sezione 3.B.1 della stessa, che sottolinea la responsabilità dei membri OCLC di garantire la consapevolezza interna ed esterna nella gestione dei dati. Nel documento si specifica, inoltre, che WorldCat ha una funzione pubblica ma non è un “bene pubblico” dal momento che l'investimento per gestirlo e mantenerlo è rilevante.

Worldcat resta un caso peculiare nel mondo dei dati bibliografici aperti: altre biblioteche e aggregatori pubblici (ad esempio: Europeana) per “licenziare” i propri dati

35 <<http://opendatacommons.org/licenses/pddl/1.0/>>.

36 La peculiarità dei dati geografici e cartografici ha fatto sì che in Europa venissero concepite licenze specifiche: ad esempio la Public Geodata License e l'Open Geodata License.

37 Cfr. Simone Aliprandi, 2011, cit.

38 La licenza ODC-BY si inserisce a metà strada tra la licenza Open Database License (ODC-ODbL), più restrittiva, e le licenze di pubblico dominio come la PDDL o la CCo.

39 *WorldCat Rights and Responsibilities for the OCLC Cooperative*: <<http://www.oclc.org/worldcat/community/record-use/policy.en.html>>.

hanno preferito la strada delle licenze di pubblico dominio⁴⁰. Di fatto, nell'ottica di un molteplici riutilizzo la scelta di adottare una licenza diversa da quelle di pubblico dominio può rivelarsi estremamente vincolante per la disseminazione dell'informazione e per garantire l'interoperabilità giuridica: «Anche le semplici e permissive licenze che richiedono la sola attribuzione sono inequivocabilmente interoperabili solo con le altre licenze più (o altrettanto) restrittive della stessa suite di licenze, ma c'è subito incertezza non appena consideriamo la possibilità, per esempio, di concedere in licenza tramite ODC-BY un dataset derivato da un insieme di dati originariamente disponibile sotto CC-BY. Infatti, in questo caso, potrebbe non essere chiaro se una licenza CC (specialmente in una versione 2.5 o precedente) garantisca le autorizzazioni necessarie per quel che riguarda il diritto *sui generis* sulle banche dati»⁴¹.

La strada delle licenze nazionali

Nonostante l'esistenza della licenza CCo e delle licenze OKFN, a partire dal 2010 diversi Paesi europei hanno sentito l'esigenza di concepire licenze utili a governare il riutilizzo delle informazioni, prevalentemente dati, prodotte nel settore pubblico (Public Sector Information).

Tra i primi a muoversi in questa direzione il Regno Unito con l'Open Government License,⁴² seguito dalla Francia con la License Ouverte⁴³ e dall'Italia con l'Italian Open Data License⁴⁴.

Sia l'Open Government License che la License Ouverte sono licenze di tipo "attribuzione". La britannica OGL include anche alcune clausole che riguardano i contenuti di tipo governativo coperti dal "Crown Copyright".

L'Italian Open Data License (IODL) viene rilasciata in versione beta ad aprile 2011 come licenza non commerciale. Ha lo scopo di promuovere la "liberazione" e valorizzazione dei dati pubblici in linea con l'articolo 52 del nuovo Codice dell'amministrazione digitale, che pone in primo piano la responsabilità delle pubbliche amministrazioni nel rendere disponibili i propri dati in modalità digitale.

Nella versione attuale, la 2.0⁴⁵, si è praticamente trasformata in una licenza di tipo "attribuzione", in quanto richiede la citazione della fonte delle informazioni e il nome del soggetto che fornisce il dato, includendo, se possibile, una copia della licen-

⁴⁰ Ad esempio: la British Library, la biblioteca del CERN di Ginevra, la biblioteca dell'università di Huddersfield e, più di recente, Europeaana. Sul sito Creative Commons si trova un elenco di casi di adozione di licenza CCo: <http://wiki.creativecommons.org/CCo_use_for_data>.

⁴¹ Federico Morando, *Interoperabilità giuridica: rendere i dati (pubblici) aperti compatibili con imprese e comunità online*, «JLIS.it», 4 (2013), n. 1, <<http://leo.cilea.it/index.php/jlis/article/view/5461>>. Sul problema delle licenze per i dati aperti si legga anche: Marco Ricolfi [et al.], *The "Licensing" of the Public Sector Information*, «Informatica e diritto», (2011), n.1/2, p.129-146 disponibile come LAPSI Position Paper n. 4 alla URL: <http://www.lapsi-project.eu/lapsifiles/LAPSI-Position_paper-Licensing-adv_draft.pdf>.

⁴² <<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/1/open-government-licence.htm>>.

⁴³ <<http://www.data.gouv.fr/Licence-Ouverte-Open-Licence>>.

⁴⁴ La Nuova Zelanda, invece, non ha concepito una licenza nazionale, ma ha adottato come soluzione di *default* la licenza CC-BY.

⁴⁵ <<http://www.dati.gov.it/iodl/2.0/>>. Sul sito InnovatoriPA è operativo un apposito gruppo di lavoro dedicato alla IODL e ai temi dell'Open Government: <<http://www.innovatoripa.it/groups/dati-aperti-open-data>>.

za o un collegamento (link) ad essa. La IODL prevede che l'utente possa liberamente: 1. consultare, estrarre, scaricare, copiare, pubblicare, distribuire e trasmettere le informazioni;

2. creare un lavoro derivato, per esempio attraverso la combinazione con altre informazioni (*mashup*), includendole in un prodotto o sviluppando un'applicazione informatica che le utilizzi come base dati.

In cambio, all'utente è chiesto di indicare la fonte delle informazioni e il nome del soggetto che fornisce il dato, includendo, se possibile, un link alla licenza. Nel 2011 è stato inaugurato il Portale dei dati aperti della PA⁴⁶ che ad ottobre 2013 raccoglie 5.776 dataset di 41 Amministrazioni.

È dubbio se e come le licenze nazionali siano tra loro interoperabili.

Marco Ricolfi (2013) sottolinea come la tendenza a rilasciare licenze di tipo nazionale possa condurre a una «Babilonia di linguaggi legali nazionali» e aggiunge: «la mia idea è che stiamo assistendo al fenomeno peculiare per cui le autorità nazionali dei singoli Stati membri credono di poter ottenere credito politico a fronte di un loro "approccio proattivo" nello sfruttamento del potenziale sociale e politico dell'informazione di settore pubblico. Il loro desiderio è che tale credito sia attribuito interamente a loro come portatori di standard nazionali piuttosto che a qualche organismo o approccio pan-europeo»⁴⁷. L'utilizzo di licenze internazionali è, senza dubbio, garanzia di una maggiore interoperabilità. Rispetto a questa esigenza di interoperabilità appare oltremodo condivisibile la petizione rivolta dalla comunità "Open data" spagnola alla Comunità Europea e al Parlamento Europeo di adottare un'unica licenza in Europa per i dati pubblici aperti. La petizione è promossa sul sito Change.org⁴⁸.

Dati aperti e Unione Europea: linee di azione e progetti

Come già accennato in precedenza, da diversi anni l'Unione Europea dimostra un interesse strategico per il tema dei dati aperti.

L'azione dell'UE si è materializzata in una riflessione concettuale prima, e in un'azione normativa poi, sui dati prodotti dalle pubbliche amministrazioni, anche in considerazione del loro potenziale economico per il mercato dell'Unione Europea⁴⁹. Nel 1999, per ricordare velocemente le tappe fondamentali di un lungo percorso iniziato in realtà nel 1989 con la pubblicazione delle Linee guida per migliorare la sinergia tra pubblico e privato,⁵⁰ l'Unione Europea pubblicava il "Libro Verde sull'informazione del settore pubblico nella società

⁴⁶ <<http://www.dati.gov.it/>>.

⁴⁷ Marco Ricolfi, *Public Sector Information as Open Data. Access, re-use and the third innovation paradigm*, in: *Extracting value from Public Sector Information: legal framework and regional policies*, a cura di Marco Ricolfi e Cristiana Sappa, Napoli-Roma: Edizioni Scientifiche Italiane, 2013, p. 23. La traduzione in italiano è mia.

⁴⁸ <<http://www.change.org/es/peticiones/say-to-neeliekroeseu-we-want-a-single-opendata-license-in-the-eu-10datalicenseeu>>.

⁴⁹ In Europa il potenziale economico dei dati prodotti dalle pubbliche amministrazioni è stato calcolato tra i 48 e i 68 miliardi di euro. Cfr. Elisa Fois – Rossana Pennanzio, *Enhancing re-use of public sector information in the European Union: Overview of practices evolved in the Member States*, in: *Extracting value from Public Sector Information: legal framework and regional policies*, a cura di Marco Ricolfi e Cristiana Sappa, Napoli-Roma: Edizioni Scientifiche Italiane, 2013, p. 43-105.

⁵⁰ Commissione Europea, *Linee guida per migliorare la sinergia tra il settore pubblico e privato nel mercato dell'informazione*, 1989.

dell'Informazione", ⁵¹ nel quale l'UE sottolineava l'importanza strategica, economica e sociale della circolazione e della trasparenza nel settore dell'informazione pubblica in Europa. Quattro anni più tardi l'UE emanava la "Direttiva 2003/98/CE del Parlamento Europeo e del Consiglio relativa al riutilizzo dell'informazione del settore pubblico" volta a favorire il riuso dei dati prodotti dalle pubbliche amministrazioni, ⁵² recentemente modificata dalla Direttiva 2013/37/CE che integra e rafforza la precedente negli scopi, estendendone, tra l'altro, l'applicazione anche ai dati e alle risorse digitalizzate da musei, archivi e biblioteche ⁵³.

Sul versante dei dati della ricerca l'azione dell'UE è stata più tardiva, ma non meno incisiva.

A luglio 2012 la Commissione Europea ha pubblicato una sua comunicazione al Consiglio e al Parlamento Europeo dal titolo "*Towards better access to scientific information: boosting the benefits of public investments in research*". Nel documento la Commissione osserva: «le discussioni sul sistema di disseminazione della ricerca scientifica si sono focalizzate tradizionalmente sull'accesso alle pubblicazioni: monografie e riviste. Tuttavia, sta diventando sempre più importante migliorare l'accesso ai dati della ricerca (risultati sperimentali, osservazioni e informazioni generate al computer) che costituiscono il presupposto per l'analisi quantitativa alla base di numerose pubblicazioni scientifiche».

Concretamente l'interesse dell'UE verso i dati della ricerca si è materializzato nel finanziamento nell'ambito del Settimo Programma Quadro di alcuni progetti volti a creare un'infrastruttura tecnologica a supporto dei dati aperti della ricerca.

OpenAIREPlus (2011-2014) è di gran lunga il progetto più significativo tra quelli finanziati dall'UE sui dati della ricerca. Nasce come costola del progetto OpenAIRE ⁵⁴ estendendone l'azione ai dati della ricerca.

Il progetto, su larga scala, riunisce 41 partner da tutta l'Europa e tre comunità di ricerca interdisciplinari. L'obiettivo principale di OpenAIREplus è la creazione di un servizio collaborativo per il cross-linking delle pubblicazioni scientifiche *peer-reviewed* e dei dataset associati. Il fine è di costruire e rendere disponibili in rete quegli oggetti digitali complessi che sono le *enhanced publications* ovvero oggetti composti da pubblicazioni associate a set di dati, immagini, video, formule, presentazioni ecc. ecc.

Nell'ambito del progetto OpenAIREPlus è stato anche realizzato Zenodo, il repository dedicato alle pubblicazioni e ai dati della ricerca. Zenodo è frutto di una stretta collaborazione tra il CERN di Ginevra e il progetto OpenAIRE, raccoglie le pubblicazioni e i dataset di progetti e comunità che non hanno un repository istituzionale o disciplinare di riferimento.

Le pubblicazioni e i dataset sono arricchiti da metadati e utilizzano come identificativo persistente per le risorse il DOI che, nella sua versione non commerciale, è mantenuto dal consorzio europeo DataCite, ⁵⁵ mentre per gli autori viene utilizza-

⁵¹ Commissione Europea, *Libro verde sull'informazione del settore pubblico nella società dell'informazione*, COM (1998) 585, <ftp://ftp.cordis.europa.eu/pub/econtent/docs/gp_it.pdf>.

⁵² <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:IT:HTML>>.

⁵³ Vedi, in particolare, i considerando 14-19, 23 e 24 e, infine, i considerando 30 e 31 sulle *partnership* pubblico-private nei progetti di digitalizzazione delle istituzioni culturali.

⁵⁴ Su OpenAIREPlus si legga: Paolo Manghi [et al.], *OpenAIREPlus: the European scholarly communication data infrastructure*, «D-Lib Magazine», 18 (2012), n. 7/8, <<http://www.dlib.org/dlib/september12/manghi/09manghi.html>>.

⁵⁵ Il DOI non commerciale <<http://www.datacite.org/>> ed EPIC <<http://www.pidconsortium.eu/>> sono attualmente i due identificativi maggiormente idonei ai dati della ricerca.

to il sistema identificativo ORCID. Zenodo comunica anche con il mondo del web sociale: dialoga, ad esempio, con Dropbox e Mendeley, due strumenti di condivisione ormai ampiamente utilizzati e consolidati tra le comunità di ricerca.

Sempre finanziato nell'ambito del Settimo Programma Quadro è il progetto EUDAT (European Data Infrastructure) lanciato nel 2011 per la durata di 36 mesi; 25 partner provenienti da 13 Paesi. EUDAT ha lo scopo di creare un'infrastruttura di servizio per la gestione dei dati della ricerca, stimolando la collaborazione e il dialogo tra centri di dati, fornitori di tecnologia e comunità di utenti. Le cinque comunità scientifiche che agiscono come partner nel progetto sono: LifeWatch (Biodiversità), ENES (Modelli climatici), EPOS (Sismologia e Vulcanologia), CLARIN (Linguistica) e VPH (Fisiologia umana).

Tra le priorità di servizio individuate dalle comunità di ricerca facenti capo al progetto rientrano: la rappresentazione e conservazione dei dati, la creazione di un'infrastruttura di servizio che consenta ai ricercatori di caricare, conservare e condividere quei dati che non vengono gestiti a livello istituzionale, la creazione di un catalogo ricercabile per tutti i dataset prodotti dalle comunità scientifiche attive in EUDAT.

Incentrato sulla condivisione, conservazione e riutilizzo dei dati è anche il progetto ODE (Open Data Exchange) lanciato nel 2010 e coordinato dal CERN. Il progetto ha raccolto una serie di storie di successo di condivisione di dati e pubblicato alcuni report tra i quali l'interessante Report on best practices for citability of data and on evolving roles in scholarly communication⁵⁶.

Ultimo progetto finanziato in Europa sul tema dei dati della ricerca il progetto RECODE (RECommendations for Open Access to Research Data in Europe)⁵⁷.

RECODE affronta il livello politico-organizzativo dell'apertura dei dati della ricerca in rete. Si propone di stilare delle raccomandazioni che servano alla definizione di un quadro politico di riferimento e alla costruzione di *policy* utili a favorire l'accesso ai dati della ricerca. A tal fine il progetto condurrà analisi su buone pratiche e studi di casi tra i vari *stakeholders* dei dati della ricerca.

A partire dal 2014 Horizon 2020, il nuovo settennale programma quadro dell'Unione Europea, ingloba un progetto pilota (Open Data Pilot) sui dati aperti, il cui scopo è facilitare la registrazione, la scoperta, l'accesso e il riutilizzo dei dati e dei metadati loro associati prodotti dai progetti di ricerca finanziati da Horizon 2020.

Biblioteche accademiche e dati della ricerca

In conclusione credo sia doverosa una breve riflessione sul rapporto tra biblioteche accademiche e dati della ricerca⁵⁸. Se è vero come è vero che la mole di dati della ricerca aumenta annualmente in modo esponenziale, che in tutta la comunità scientifica internazionale e tra gli enti finanziatori sta crescendo il livello di attenzione verso i dati della ricerca, che sempre più di frequente i dati grezzi affiancano e completano le pubblicazioni scientifiche, ne consegue che le biblioteche accademiche, per venire incontro alle mutanti esigenze della comunità scientifica, devono ritagliarsi un ruolo nell'indicizzazione, gestione, conservazione e distribuzione in rete dei dati della ricerca.

⁵⁶ Accessibile alla URL: <<http://www.alliancepermanentaccess.org/index.php/community/current-projects/ode/outputs/>>.

⁵⁷ <<http://recodeproject.eu/>>.

⁵⁸ Sull'argomento – si legga: Carol Tenopir – Robert J. Sandusk – Suzie Allard – Ben Birch, *Academic librarians and research data services: preparation and attitudes*, «IFLA Journal», 39 (2013), n. 1, p.70-78.

Nel mondo anglosassone, già nel 2008, in un report finanziato dal Joint for Information Systems Committee (JISC) e realizzato dalla Key Perspective, vengono distinti i ruoli e le competenze necessarie alla gestione dei dati della ricerca: il *data librarian*, il *data creator*, il *data scientist* e il *data manager*⁵⁹.

Il ruolo del *data librarian* trae chiaramente origine dal contesto dei depositi digitali (*repository*); la figura professionale appare riferita in modo specifico alla gestione e conservazione dei dati, ma concettualmente è parte integrante di una figura più ampia e trasversale che abbraccia le diverse comunità professionali che si occupano di trattamento e conservazione del patrimonio culturale e scientifico: il *digital curator*⁶⁰. Il rapporto tra *data librarian* e *digital curator* risulta, in realtà, ancora incerto, i contorni delle due figure si sovrappongono, così come le competenze. Tra i due profili il *data librarian* sembra essere più vicino al mondo della ricerca scientifica, il *digital curator* al mondo degli archivi, biblioteche e musei (*Cultural Heritage*).

Parallelamente stanno nascendo in ambito anglo-americano nuovi servizi bibliotecari legati alla gestione dei dati della ricerca: ad esempio, le biblioteche Sheridan della John Hopkins University offrono un servizio a pagamento per la gestione dei dati, mentre la Bodleian Library dell'Università di Oxford sta sviluppando un sistema di conservazione a tariffe differenziate rispetto alla tipologia di dati da gestire e conservare⁶¹.

Per le biblioteche la gestione e conservazione dei dati pone l'ennesima serie di sfide: di formazione, di competenze e di riconoscimento delle figure professionali emergenti; di potenziamento dell'infrastruttura tecnologica - quella esistente è gestita in modo prevalente dalle biblioteche accademiche e di ricerca e necessita di essere potenziata e adattata alle peculiarità dei dataset;⁶² di cambiamento culturale e organizzativo; di finanziamenti; di pianificazione progettuale; di collaborazione tra *stakeholders* diversi; di *policy* coerenti e cogenti⁶³.

L'Open Data Pilot del programma Horizon 2020 costituisce un'opportunità unica per stimolare tra i ricercatori una cultura della condivisione dei dati. L'importanza strategica del programma pilota è stata sottolineata anche dal recente comunicato congiunto redatto a luglio 2013 da OpenAIRE, LIBER (Ligue des Bibliothèques

59 Alma Swan – Sheridan Brown, *The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs*, Report to the JISC, Key Perspectives Ltd, July 2008, <<http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreport.aspx>>.

60 Sul *digital curator* in italiano si leggano: Mariella Guercio, *Le discipline del documento e l'innovazione tecnologica nelle iniziative di formazione degli archivisti*, «DigItalia», 6 (2011) n. 1, <<http://digitalia.sbn.it/article/view/204>>; Laura Testoni, *Digital curation and content curation: due risposte alla complessità dell'infosfera digitale che ci circonda, due sfide per i bibliotecari*, «Bibliotime», 16 (2013), n.1, <<http://www.aib.it/aib/sezioni/emr/bibtime/num-xvi-1/testoni.htm>>; Maria Cassella, *Il digital curator*, «Biblioteche oggi», 31 (2013), n. 6, p. 3-10.

61 Cfr. il post del 19 giugno 2013 *Le biblioteche alle prese con i Big Data* sul blog "Cultura digitale" di Fabio di Gianmarco: <<http://www.culturadigitale.it/wp/linked-data/148/le-biblioteche-alle-prese-con-il-E2%80%9Cbig-data%E2%80%9D/>>.

62 Non mancano i registri di *repository* dedicati ai dati grezzi. Registri di *data repository* sono: re3data <<http://www.re3data.org/>>, mantenuto dalla German Research Foundation, e Databib <<http://databib.org/>>.

63 Il progetto europeo MedOANET ha tra i suoi obiettivi finali la pubblicazione di una *policy* sugli Open Research Data.

Européennes de Recherche) e COAR (Confederation of Open Access Repositories) in risposta ad una consultazione pubblica della Commissione Europea sugli Open data⁶⁴. Il comunicato assegna alle biblioteche, insieme a editori, ricercatori, centri di ricerca ecc., un ruolo chiave nella gestione dei dati. Le biblioteche provvederanno alla descrizione dei dataset prodotti dai gruppi di ricerca locali, che dovranno essere descritti utilizzando schemi standard di metadati⁶⁵ anche quando non direttamente accessibili in rete. I dataset dovranno essere depositati in *repository* certificati, associati alle pubblicazioni e dotati di identificativi persistenti.

In Italia esiste ancora una scarsa consapevolezza del ruolo che le biblioteche potrebbero ritagliarsi nel dominio dei dati della ricerca; le comunità scientifiche restano gli attori principali nella gestione dei dati della ricerca. Tuttavia dall'Open Access agli Open Data il passaggio è naturale, se non obbligato, soprattutto per quei bibliotecari che lavorano nell'ambito della comunicazione scientifica e sono a stretto contatto con le comunità di ricerca: il fine ultimo è di inglobare nella gestione dei *repository* la "cura" dei dati della ricerca, soprattutto in quelle realtà dove non esistono strategie complementari che siano in grado di garantire altrettanto efficacemente interoperabilità e conservazione a lungo termine dei dati.

ABSTRACT

AIB studi, DOI 10.2426/aibstudi-9131, vol. 53 n. 3 (settembre/dicembre 2013), p. 223-238.

MARIA CASSELLA, Università degli studi di Torino, Sistema bibliotecario di ateneo, via Verdi 8, 10124 Torino, e-mail maria.cassella@unito.it.

Dati aperti e ricerca scientifica: aspetti gestionali e normativi nel contesto dell'e-science

Nel campo della gestione dei dati della ricerca si stanno aprendo nuovi scenari per le comunità scientifiche, nonché per le biblioteche accademiche e degli enti di ricerca. Accanto a questi nuovi scenari, tuttavia, si vanno delineando anche nuove sfide e interrogativi. Partendo da alcune definizioni di "dati aperti", l'autrice descrive i principali vantaggi dell'accesso in rete ai dati della ricerca, discutendo diffusamente degli aspetti normativi della questione, di dove e quando i dati possono essere messi liberamente a disposizione in rete, a quali condizioni, e con quali licenze. L'ultimo capitolo è dedicato al tema della gestione – indicizzazione, conservazione e distribuzione in rete – dei dati della ricerca come nuovo possibile paradigma di attività per le biblioteche accademiche e di ricerca.

Open data and scientific research: the organizational and regulatory framework of e-science

The dissemination of research data opened an entirely new framework for scientific communities as well as for academic and research libraries. Besides, it generated new issues and questions. Starting with some definitions of "open data", the author outlines the benefits of open access to research data and analyze its

64 Horizon 2020: outline of a Pilot for Open Research Data. Joint statement by OpenAIRE, LIBER and COAR, 3 July 2013, <http://www.openaire.eu/en/about-openaire/publications-presentations/publications/doc_details/585-horizon2020opendatapilot20130703final>.

65 Sui diversi schemi di metadati utilizzati per descrivere i *research data* segnalo la pagina web mantenuta dal Digital Curation Centre: <<http://www.dcc.ac.uk/resources/metadata-standards>>.

regulatory framework: where, when, under which conditions and with what kind of license can research data be made available online? The last part of the article illustrates how research data organization – indexing, management, preservation and access – can become a new professional paradigm for academic and research librarians.